Peer assessment of teamwork in group projects: Evaluation of a rubric

Peer assessment of teamwork in group projects: Evaluation of a rubric

Brett Vaughan
University of Melbourne, Australia
Jacqui Yoxall and Sandra Grace
Southern Cross University, Australia

Effective teamwork is integral to the functioning of many professions. Employers desire work-ready graduates who have developed the so-called 'soft skills' involved in teamwork. One way for students to practise and gain feedback on their team skills is through team-based assessment tasks. The TeamUP Rubric was developed to teach and assess five domains of teamwork skills through peer feedback. Initial evaluation of the five domains and use of the TeamUP with health and arts students has provided initial support for its utility as a measure of teamwork. In the current study the scale structure of each of the five domains was evaluated using Mokken scale analysis (MSA). MSA is a non-parametric item response theory approach that evaluates the internal structure of an evaluation tool. One-hundred and seventy-seven primary education students were recruited for the study. All participants were undertaking a teamwork assignment in a unit of study in their education degree at an Australian regional university. They completed the TeamUP Rubric assessment for themselves and for each of the students in their team. Each TeamUP domain met the requirements for a Mokken scale. The current study supports the validity of the scores derived from the TeamUP Rubric and further supports its use as a tool for teaching and assessing teamwork.

Introduction

The literature suggests that a team comprises a group of individuals who are adaptable, and undertake roles and tasks that are interrelated in order to achieve a common outcome (Salas & Cannon-Bowers, 2001; Varela & Mead, 2018). To achieve this outcome, each member of the team requires knowledge of specific tasks (both for themselves and other members of the team), skills to be able to evaluate their own performance and that of other team members, and their attitude towards engaging in teamwork. Importantly, such knowledge, skills and attitudes need to be transferable to other contexts and teams (Baker, Day & Salas, 2006; Sinche et al., 2017).

Teamwork skills are an important workplace competency (Coll & Zegwaard, 2006; Quek, 2005) and the ability to work in a team and work with others, including in virtual team environments (Horvath & Tobin, 2001), is often considered a fundamental employability skill for graduates from higher education programs who aspire to professional work (Betta, 2016). The *Australian Qualifications Framework 2nd Edition* (2013) stated that at bachelor degree level graduates will demonstrate 'well developed … communication skills' to address and find solutions to complex problems and to '… transmit knowledge, skills and ideas to others' (p,47). Furthermore, the *Framework* states that 'Graduates of a Bachelor Degree will demonstrate the application of knowledge and skills ... with responsibility and accountability for their own learning and professional practice and in

collaboration with others within broad parameters' (p48). Together, these suggest that university graduates should have the ability and disposition for working in teams (Ruiz Ulloa & Adams, 2004).

However, according to Leggat (2007), focusing on individual skill development for members of a team may not fully encompass competencies required for effective teamwork. Leggat (2007) described competencies for effective teamwork in healthcare management teams - working collaboratively, quality outcomes and a commitment to the institution – as distinct from competencies for individual skill development, accountability and achievement in teams. This notion was supported by Mickan and Rodger (2000) who summarised 18 characteristics of effective teams in the literature across organisational, team-based contributions and the influence of the individual. Shared mental models also appear to underpin effective teams, particularly where there is organisational support (Druskat & Pescosolido, 2002).

Effective teams provide both team-based outcomes and benefits for individuals participating in teams. Teamwork training promotes an individual's increased understanding of teamwork competencies and improved competence in planning and coordinatiing, communication skills, and problem□solving in a team environment (Ellis, Bell, Ployhart, Hollenbeck & Ilgen, 2005). Further, teamwork is also reported to be positively associated with a range of psychological variables including workplace satisfaction, autonomy and also identifying oneself as part of a team (Rousseau, Aubé & Savoie, 2006; Taylor et al., 2019). Conversely, poor teamwork has been associated with anxiety and absenteeism in workplace settings (Asfaw, Chang & Ray, 2014).

Several approaches have been taken to evaluate teamwork skills. For example, a variety of questionnaires have been developed to measure teamwork competencies and perceptions using observation or self-report and/or information obtained from learning analytics systems (e.g. Aguado, Rico, Sánchez-Manzanares & Salas, 2014; Baker, Amodeo, Krokos, Slonim & Herrera, 2010; Cooke et al., 2003; Fidalgo-Blanco, Sein-Echaluce, García-Peñalvo & Conde, 2015; Keebler et al., 2014; Valentine, Nembhard, & Edmondson, 2015; Varela & Mead, 2018). Most of these measures attempt to tap into communication, coordination, respect and use of members' expertise in group work. However, the evaluation of the complex construct of teamwork, or the constructs that together contribute to teamwork is complex at best. However, many evaluation tools established to date lack a sound theoretical basis and/or psychometric properties (Varela & Mead, 2018), rendering many measures unsuitable for use in teamwork training, feedback, decision-making and remediation (Rosen et al., 2008). Others are context specific, thereby limiting generalisability to a wide variety of disciplines or circumstances.

In the teaching and learning context, peer assessment is widely used as a strategy to assist in the development of reflective and lifelong learners (Dochy, Segers & Sluijsmans, 1999). There appears to be value in the use of peer assessment for evaluating "…academic products and processes, rather than professional practice …" (Falchikov & Goldfinch, 2000, p. 287) and when global assessments are used rather than assessments of individuals. In these cases peer marks were found to closely align with faculty assessments.

Peer assessment may also be strengthened by the inclusion of peer feedback, particularly before a peer assessment mark is provided (Liu & Carless, 2006). For such peer assessment to be a valid and reliable, peers must be trained for the assessment (Van Zundert, Sluijsmans & Van Merriënboer, 2010).

In Australia, Hastie et al. (2014) developed the peer assessment rubric 'TeamUP' and associated educational material to teach and evaluate teamwork skills, initially for undergraduate midwifery students. The rubric was developed by drawing on the theoretical teamwork literature, focusing on five domains: project planning skills; fostering a team climate; facilitating contributions of others; managing conflict; and, contributing to a team project. In subsequent work, Parratt et al. (2016) used the Delphi method to explore the validity of the TeamUP Rubric. This work resulted in the maintenance of the five-domain structure, with some some modifications to the wording of some items. According to Parratt et al. (2016), the TeamUP Rubric can be used '… to teach and assess teamwork skills with confidence …', and that educators '… are using a well-validated tool to do so' (p. 84). Additional work by Britton, Simper, Leger and Stephenson (2017) utilising the TeamUP Rubric with theatre history and literature students in Canada provided psychometric support for the structure and internal consistency of the rubric. Although quantitative evidence to support the 'well validated tool' assertion, beyond the Britton et al. (2017) study, is limited, there is a sound theoretical framework underpinning the rubric, addressing the concern of Varela and Mead (2018) described previously. The aim of the present study was to continue the development of the TeamUP Rubric (Hastie et al., 2014) by evaluating the internal structure of the measure, using an item response theory approach, in a pre-service education student cohort.

Method

The study was approved the by the Southern Cross University Human Research Ethics Committee (ECN 13-037).

Participants

A total of 177 undergraduate students enrolled in a first-year education unit (Curriculum and Programming) at an Australian regional university and undertaking a group work project as part of their program, accepted an invitation to participate in the study. Participants were involved in a compulsory group work project, which was conducted over a period of 12 weeks and carried 50% of the overall assessment weighting for the unit of study. Students were randomly allocated to groups by the unit of study coordinator. Projects were organised to ensure that students had allocated time, both face to face and online, to work together. Students' previous experiences of teamwork varied according to their age and life experience. Each participant completed a hard copy version of the TeamUP Rubric as a self-assessment, and for each of the students in their team. Assessments were completed in such a way as to conceal the identity of the individual assessors from those they were assessing, but not from the academic staff member managing the unit of study. Overall there were 576 cases and after data cleaning (removal of cases where data were systematically missing), 527 were available for analysis.

Assessment tool

The TeamUP Rubric was developed by Hastie et al. (2014) and Parratt et al. (2016) to peer assess teamwork skills. The latter authors used the Delphi method to reach consensus for items that assess teamwork skills across five domains: *planning*; *environment*; *facilitation*; *conflict management*; and, *individual contribution*. Content validation index values in the Delphi study ranged from 77% to 93% for the five domains. In the current study, students were required to complete the TeamUP Rubric in *Qualtrics* (USA) and identify both themselves and the students they were rating. Responses to each item were recorded on a 7-point Likert-type scale with an additional 'not applicable' option (Appendix 1).

Data analysis

Data were extracted from *Qualtrics* and then exported to *SPSS* for cleaning. Next, data were exported to *R* (R Core Team, 2016) for analysis. Missing data (data points missing at random) were imputed using the two-way imputation in the *TestDataImputation* package. This methodology has been advocated for use with Mokken scaling (Sijtsma & Van der Ark, 2003; Van Ginkel, Van der Ark, & Sijtsma, 2007). Each of the five TeamUP Rubric domains were evaluated separately. Descriptive statistics (median, range) were generated using the *psych* package (Revelle, 2016) and a Mokken scale analysis (MSA) using the Mokken package (Van der Ark, 2012).

Mokken scale analysis (MSA)

MSA is a non-parametric item response theory technique used to evaluate the dimensionality and structure of a scale. A Mokken scale for dichotomous responses is formed when the data meet the three assumptions underlying this approach (Stochl, Jones, & Croudace, 2012): unidimensionality (all items measure the same underlying construct); local item independence (response to one item should not systematically influence responses to another item); and, monotonicity (probability of selecting a particular response should occur in a non-decreasing manner). As the TeamUP Rubric items are polytomous, an additional assumption of non-intersection (invariant item ordering) was included. This assumption is met when the item response curves do not overlap.

For the Mokken scale analysis, the steps described by both Wind (2017) and Siijtsma and van der Ark (2017) were followed:

1. Identify missing data and outliers.
2. Evaluate the scalability of the measure (does the measure evaluate a single construct?)
3. Local independence (is an item systematically influenced by responses to another item?)
4. Monotonicity (as the level of the construct goes up, does the responses to an individual item follow this pattern?)
5. Invariant item ordering (can the items be ordered to represent the construct being measured?)
6. Scale reliability (after addressing issues identified in steps 2-5 above).

Mokken coefficients are described in the Results: H*i* refers to the individual item coefficient, H*ij* refers to the item pair coefficient, and H^T is the coefficient demonstrating the strength of the item ordering, that is, high levels of the construct are represented by one item over another.

Reliability estimation

Internal structure was evaluated using McDonald's omega total (ωt) and omega hierarchical (ωh) (Revelle & Zinbarg, 2009; Zinbarg, Revelle, Yovel & Li, 2005; Zinbarg, Yovel, Revelle & McDonald, 2006). ωh values greater than 0.50 (Revelle, 1979) support the calculation of a total score and ωt values greater than 0.70 provide support for the internal structure of the questionnaire.

Results

One hundred and seventy-seven (177) students were enrolled in a first-year education unit (Curriculum and Programming). As part of their group project they were required to use the TeamUp Rubric to rate their own performance and that of peer team members. Overall 574 entries were collected, and after data cleaning 527 entries were available for analysis. Descriptive statistics and Mokken item coefficients (H*i*) for each TeamUP item are presented in Table 1 as evidence of the internal structure and dimensionality of the TeamUP Rubric using Mokken scale analysis. Results for each domain are described below from the perspective of local dependence (i.e. a response to one item in the scale is significantly influenced by a response on another) and invariant item ordering (IIO) (i.e. when the item order varies regardless of the level of the underlying construct).

Table 1: Descriptive statistics and scale coefficients for the TeamUP Rubric on initial analysis

Domain	Median	Range	H_i coefficient (± std. error)
Domain 1 – Planning			0.867 (0.014)
a. Electing and supporting a project manager	2	1-7	0.845 (0.018)
b. Defining and agreeing on team goals and objectives	1	1-7	0.870 (0.016)
c. Defining and agreeing on quality standards for each part of the plan	1	1-7	0.872 (0.014)
d. Contributing to the development of the plan	1	1-7	0.877 (0.015)
e. Setting and agreeing realistic timeframes for each part of the plan	1	1-7	0.852 (0.019)
f. Participating in role allocations based on individual skills and learning needs	1	1-7	0.877 (0.015)
g. Willingly taking on a team role that can be completed on time to a quality standard	1	1-7	0.881 (0.016)
Domain 2 - Environment			0.824 (0.021)
a. Exhibiting an open, gentle, polite and friendly manner	1		0.807 (0.023)
b. Demonstrating self-awareness and emotional regulation	1	1-7	0.821 (0.020)

c. Demonstrating sensitive awareness of the feelings of others (including interpreting body language)	1	1-7	0.830 (0.021)
d. Actively contributing to team discussions	1	1-7	0.777 (0.026)
e. Cooperating with others to achieve project goals	1	1-7	0.839 (0.022)
f. Following up with others when there is concern about their feelings or contribution	1	1-7	0.823 (0.024)
g. Showing respect for the contributions of others (even if in disagreement)	1	1-7	0.833 (0.020)
h. Expressing genuine gratitude and praise generously	1	1-7	0.830 (0.021)
Domain 3 - Facilitating the contributions of others			0.774 (0.023)
a. Taking turns at leading/coordinating a team meeting (including agenda preparation)	2	1-7	0.771 (0.026)
b. Taking turns at keeping and distributing brief meeting minutes (with action items and deadlines)	2	1-7	0.747 (0.034)
c. Leading and/or participating in teambuilding processes	2	1-7	0.801 (0.022)
d. Establishing and honouring team ground rules	1	1-7	0.787 (0.023)
e. Ensuring that decisions are made in a timely manner	1	1-7	0.781 (0.023)
f. Listening attentively without interrupting and raising hand to speak	1	1-7	0.743 (0.026)
g. Participating in consensus-building decision-making	1	1-7	0.781 (0.026)
h. Inviting other team members to contribute	1	1-6	0.776 (0.023)
Domain 4 – Managing conflict			0.841 (0.019)
a. Being appropriately assertive: neither dominating, submissive, nor passive aggressive	1	1-7	0.787 (0.026)
b. Expressing concerns with team/team members in a constructive manner	1	1-6	0.849 (0.019)
c. Minimising unnecessary conflict by project planning and management	1	1-7	0.868 (0.016)
d. Completing assigned responsibilities on time	1	1-7	0.800 (0.030)
e. Participating in the team conflict transformation processes	1	1-6	0.874 (0.016)
f. Assisting the team to stay focused on the overall team goal	1	1-7	0.828 (0.025)
g. Approaching conflict with the aim to de-escalate	1	1-7	0.867 (0.018)
h. Being open to receiving and reflecting upon criticism of self	1	1-7	0.837 (0.020)
i. Challenging team processes that are not conducive to the achievement of team goals	1	1-6	0.866 (0.017)
Domain 5 – Contributing to Team Project			0.828 (0.019)
a. Demonstrating sufficient technological skills	1	1-6	0.733 (0.041)
b. Demonstrating relevant content knowledge	1	1-6	0.830 (0.021)
c. Adhering to academic standards for writing	1	1-7	0.839 (0.023)
d. Submitting assigned work at the agreed quality standard	1	1-7	0.856 (0.020)
e. Submitting assigned work within the agreed timeframe	1	1-7	0.815 (0.025)
f. Appropriately critiquing the work of others	1	1-7	0.832 (0.023)
g. Working to integrate the output of team members into the project	1	1-7	0.875 (0.015)
h. Evaluating the quality of the whole project and making needed changes	1	1-7	0.833 (0.019)

Domain 1 (Planning)

Domain 1 (Planning) demonstrated high H_i coefficients (>0.83) and positive H_{ij} coefficients. One locally dependent item pair was flagged (Items 1 and 2). Item 1 'Electing and supporting a project manager' was removed due to its lower H*i* value and the domain no longer displayed any locally dependent items. After removal of Item 1, the domain items did not display any violation of monotonicity nor invariant item ordering. A H^T value of 0.14 was demonstrated suggesting weak ordering of the items measuring the Planning construct.

Domain 2 (Environment)

Domain 2 (Environment) demonstrated high H_i coefficients (>0.76), positive H_{ij} coefficients, no local item dependence and no monotonicity violations. Issues with invariant item ordering (IIO) were identified for Item 3 'Demonstrating sensitive awareness of the feelings of others (including interpreting body language)', Item 4 'Actively contributing to team discussions' and Item 5 'Cooperating with others to achieve project goals'. None of these violations was statistically significant, therefore the items were retained. The H^T value of 0.12 suggested the item ordering was weak.

Domain 3 (Facilitating the contribution of others)

Domain 3 'Facilitating the contribution of others' demonstrated high H_i coefficients (>0.77) and positive H_{ij} coefficients. Item 6 '[Listening attentively without interrupting and raising hand to speak' demonstrated local dependence with Items 1-3 and was removed. Following removal of Item 6, Item 1 was also removed due to local dependence with Items 5 and 7. Item 2 then demonstrated local dependence with Item 5 and was removed. Finally, Item pair 5 and 8 demonstrated local independence and Item 8 was removed. With these items removed, coefficient H was 0.812 (0.023) with no violation of monotonicity, no invariant item ordering and a H^T value of 0.03.

Domain 4 (Managing conflict)

Domain 4 'Managing conflict' demonstrated high H_i coefficients (>0.77) and positive H_{ij} coefficients. Items 1 and 3 demonstrated local dependence and Item 1 was removed. Item 4 then demonstrated local independence with Items 2 and 7, in addition to Item pair 6 and 8. Item 4 was removed and this resolved the local independence for item pair 6 and 8. Following removal of Items 1 and 4, coefficient H for the domain was 0.874 (0.016), however significant violations of IIO were identified. Analysis suggested the removal of Items 2 and 8 and once removed the coefficient H value was 0.888 (0.016), no violation of monotonicity or invariant item ordering was identified and the H^T value was 0.007.

Domain 5 (Contributing to team project)

Domain 5 'Contributing to team project' demonstrated high H_i coefficients (>0.73) and positive H_{ij} coefficients. Local dependence was identified between Item pairs 6 and 7, and 7 and 8; Item 7 was subsequently removed. Following removal, the coefficient H value

was 0.812(0.021), no violation of monotonicity or invariant item ordering were identified and the H^T value was 0.31.

Reliability estimates

The final TeamUp Rubric with five domains is found at Table 2. Reliability estimations (McDonald's omega (ω)) were calculated for each of the TeamUP Rubric domains and coefficients are presented in Table 3. The ωh values suggest that the underlying latent construct for each domain is accounted for in the total domain score. The reliability estimations provide evidence to support the calculation of a total score for each domain.

Table 2: Final TeamUP Rubric consisting of five domains

Domain 1: Planning	1	Defining and agreeing on team goals and objectives
	2	Defining and agreeing on quality standards for each part of the plan
	3	Contributing to the development of the plan
	4	Setting and agreeing realistic timeframes for each part of the plan
	5	Participating in role allocations based on individual skills and learning needs
	6	Willingly taking on a team role that can be completed on time to a quality standard
Domain 2: Environment	1	Exhibiting an open, gentle, polite and friendly manner
	2	Demonstrating self-awareness and emotional regulation
	3	Demonstrating sensitive awareness of the feelings of others (including interpreting body language)
	4	Actively contributing to team discussions
	5	Cooperating with others to achieve project goals
	6	Following up with others when there is concern about their feelings or contribution
	7	Showing respect for the contributions of others (even if in disagreement)
	8	Expressing genuine gratitude and praise generously
Domain 3: Facilitating contributions of others	1	Leading and/or participating in team building processes
	2	Establishing and honouring team ground rules
	3	Ensuring that decisions are made in a timely manner
	4	Participating in consensus-building decision-making
Domain 4: Managing conflict	1	Expressing concerns with team/team members in a constructive manner
	2	Minimising unnecessary conflict by project planning and management
	3	Participating in the team conflict transformation processes
	4	Assisting the team to stay focused on the overall team goal
	5	Approaching conflict with the aim to de-escalate
	6	Being open to receiving and reflecting upon criticism of self
	7	Challenging team processes not conducive to the achievement of team goals
Domain 5: Contributing to team project	1	Demonstrating sufficient technological skills
	2	Demonstrating relevant content knowledge
	3	Adhering to academic standards for writing
	4	Submitting assigned work at the agreed quality standard
	5	Submitting assigned work within the agreed timeframe
	6	Appropriately critiquing the work of others
	7	Evaluating the quality of the whole project and making needed changes

Table 3: Reliability estimates for the five TeamUP domains subscales using McDonald's omega (ω) and Mokken's rho

Scale	Omega total (ωt)	Omega hierarchal (ωh)	Rho
Domain 1 – Project planning skills	0.97	0.92	0.97
Domain 2 – Fostering a team climate	0.98	0.89	0.96
Domain 3 – Facilitating the contribution of others	0.95	0.89	0.93
Domain 4 – Managing conflict	0.98	0.92	0.97
Domain 5 – Contributing to team project	0.97	0.92	0.96

The Unit Assessor of the education unit reported a number of negative comments about the use of the TeamUP Rubric by students. Comments were mostly linked to students' dislike of team assignments and this was a high value task. They found the document tedious and felt that the measure did not allow for identification of team members who did their job well.

Discussion

Developing teamwork skills must be an integral part of university learning and assessment if new graduates are to meet industry expectations (Hastie, 2018; Leggat, 2007; Mickan & Rodger, 2000; Ruiz Ulloa & Adams, 2004) and group assignments are commonly used for this purpose. However, students dislike the allocation of the same grade to all members of the team, regardless of individual contribution. In fact, LaBeouf, Griffith, and Roberts (2016) reported grading as the most important issue for students undertaking group assignments. This was reflected in the Unit Assessor's feedback from students in our study. Most commonly, teamwork skills are assessed using survey instruments (Valentine et al., 2015), however, many such instruments have inadequate psychometric properties or are context dependent, and therefore not generalisable to other contexts and disciplines (Varela and Mead 2018). Peer-assessment rubrics, like the TeamUP Rubric, focus on self-evaluation and evaluation of the performance of others in the team (Baker et al., 2006). The ongoing development and validation of the TeamUP Rubric as a learning tool for students and their teams accords with the requirement of universities to develop teamwork skills in their students (Australian Qualifications Framework Council, 2013) and also provides a tool to assist with the assessment of teamwork learning outcomes within a curriculum (Britton et al., 2017).

The present study evaluated the internal structure and dimensionality of the TeamUP Rubric using Mokken scale analysis. Each TeamUP domains was analysed independently to evaluate whether the requirements of a Mokken scale were met. Local dependence was the major measurement issue identified in the TeamUp Rubric. That is, a response to one item in the scale was significantly influenced by the response on another. For Domain 1 (Planning), Items 1 and 2 were locally dependent. Responses to Item 2 'Defining and agreeing on team goals and objectives' likely subsumed those of Item 1 'Electing and supporting a project manager' whereby a project manager was to be identified as part of setting the team goals and objectives. Domain 3 required the greatest number of items to

be removed due to local dependence. Item 6 'Listening attentively without interrupting and raising hand to speak' was removed as it was dependent with other items related to the process aspects of teamwork 'Leading a meeting, Minute taking, Team building' all of which require active listening. Item 1 'Leading a meeting' was subsequently removed due to dependence with Items 5 'Timely decision making' and 7 'Consensus decision-making' suggesting leadership was influenced by both timely and consensus decisions. Item 8 'Inviting contributions' and Item 5 'Timely decisions' were also locally dependent suggesting that timely decision-making required contributions from other team members. In Domain 4 Items 1 'Appropriately assertive' and 8 'Minimising conflict' were locally dependent and the more 'positive' of the two items (Item 8) was retained. Unsurprisingly, Item 4 'Completing assigned responsibilities on time' was locally dependent with multiple items given that conflict can arise when team members do not complete their assigned tasks in a timely manner. For Domain 5 Item 7 'Integration of outputs' was locally dependent with both Item 6 'Critique of work' and Item 8 'Quality of project output'. This result suggests that integration was a requirement of both critique (bringing the work together in order to critique it as a team) and project quality (bringing the work together in order to evaluate its quality). By addressing the issue of local dependence, the validity of the information derived from each TeamUP Rubric domain was strengthened – items were removed that were systematically influenced by other responses or influencing responses to other items, providing a more accurate representation of the construct (teamwork) being measured.

Invariant item ordering (IIO) was identified in Domain 4 for both Item 2 'Expressing concerns with team/team members in a constructive manner' and Item 8 'Being open to receiving and reflecting upon criticism of self'. IIO occurs when the item characteristic curves cross over with other items on the same scale. How these two items are scored by an individual is not consistent with the underlying construct. For example, Item pair 2 and 6 demonstrates that participants with a lower total score for Domain 4 (Managing conflict) were more likely to select higher responses to Item 2 than for Item 6, and this was corrected with higher Domain 4 total scores. Essentially, Item 2 did not elicit a response in the same way as the other items in the Domain, suggesting it was measuring something other than the underlying construct.

Results of the study suggest that each of the domains are scalable but no conclusion can be drawn regarding the ordering of the items within each domain due to the low H^T values. It has been suggested that values less than 0.3 are uninterpretable with regard to item ordering within each domain (Sijtsma & van der Ark, 2017). This result suggests that one item in a domain is not more likely to be selected than another, given increasing values of the underlying construct. This is unlikely to be problematic for the TeamUP Rubric because it would not be expected that one item represents more of the latent construct than another.

The calculation of a total score for each of the five domains is possible and likely represents the construct of each domain. The reliability estimations support these calculations with values for both McDonald's *omega* and Mokken's *rho* over 0.90. Although there is literature that suggest mid-points on Likert-type scales may be problematic,

participants in this study used the full Likert scale range for each TeamUp item, suggesting utility of the mid-point.

This study has provided further evidence of validity of the TeamUP Rubric (Parratt et al., 2016). Data presented here suggest that teamwork skills may be generalisable across professions and work environments and that the TeamUP Rubric may be a useful tool to assess these skills, given its exploration now with both midwifery and education students. This is consistent with the results of a survey of 8099 PhD science students conducted by Sinche et al. (2017), wherein the authors identified 15 transferrable skills, including teamwork. The initial validation of the TeamUP Rubric was in a health profession student population and the subsequent investigation here with education students, alongside the work by Britton et al. (2017), provides further evidence of construct validity and utility in other student populations.

Limitations

There are a number of limitations in the current work. The results of the study may not be generalisable beyond the single unit of study, or single institution in the current work. Whether the TeamUP is context-specific will require further investigation. Students may have rated themselves and their peers high (or potentially low), given the summative nature of the task, and this could bias the outcome of the study. It is also possible that students reflected their experience of the unit of study as a whole through the TeamUP Rubric, although this appears to be less likely, given that the statistical outcomes support the measurement of a single construct, consistent with Britton et al. (2017). Nadler, Weston and Voyles (2015) suggested that the mid-point of a Likert-type scale be defined if it is to be used; however this was not done in the current work. As such, the use of this midpoint for a TeamUP Rubric item is possibly influenced by the anchoring comments either side of it; the Likert-type scale is not being used as intended. However, a strength of including a midpoint is that the student may have truly been 'indifferent' to the performance of a peer with respect to a particular item (Chyung, Roberts, Swanson & Hankinson, 2017).

Future research

Further work is required to evaluate the TeamUP Rubric in other student populations, to explore divergent and convergent validity with other assessments of teamwork, and to potentially evaluate the utility of the short-form rubric developed by Britton et al. (2017) in other populations. Such a version could prove useful as a formative assessment tool. Qualitative investigations could also explore the students' conceptions of each of the TeamUP items and the Likert-type scale used. There is also a need to explore predictive validity of the measure and in particular to explore whether TeamUP Rubric results reliably predict teamwork performance in other contexts, including in their professional life. Until such evidence is produced, students, lecturers and future employers may remain skeptical of the usefulness of peer assessment and group projects in developing teamwork skills in students.

Conclusion

Evaluation of teamwork skills may help to facilitate the development of peer assessment capacity and teamwork skills themselves. The TeamUP Rubric may provide a framework for the development, evaluation and assessment of teamwork skills. The present study provides support for the internal structure of the TeamUP Rubric through the lens of a modern evidence-informed test theory. The scores derived from each of the five TeamUP domains can be said to reflect the latent construct that is the domain. This study supports the validity argument for the TeamUP Rubric for use in different populations and supports the use of total score for each domain, thereby providing further evidence for its usefulness in teaching and assessing teamwork skills.

References

Aguado, D., Rico, R., Sánchez-Manzanares, M. & Salas, E. (2014). Teamwork Competency Test (TWCT): A step forward on measuring teamwork competencies. *Group Dynamics: Theory, Research, and Practice,* 18(2), 101-121. https://doi.org/10.1037/a0036098

Asfaw, A. G., Chang, C. C. & Ray, T. K. (2014). Workplace mistreatment and sickness absenteeism from work: Results from the 2010 National Health Interview survey. *American Journal of Industrial Medicine*, 57(2), 202-213. https://doi.org/10.1002/ajim.22273

Australian Qualifications Framework Council (2013). *The AQF second edition January 2013*. https://www.aqf.edu.au/aqf-second-edition-january-2013

Baker, D. P., Amodeo, A. M., Krokos, K. J., Slonim, A. & Herrera, H. (2010). Assessing teamwork attitudes in healthcare: Development of the TeamSTEPPS teamwork attitudes questionnaire. BMJ Quality and Safety, 19(6). https://qualitysafety.bmj.com/content/19/6/e49

Baker, D. P., Day, R. & Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health Services Research*, 41(4p2), 1576-1598. https://doi.org/10.1111/j.1475-6773.2006.00566.x

Betta, M. (2016). Self and others in team-based learning: Acquiring teamwork skills for business. *Journal of Education for Business*, 91(2), 69-74. https://doi.org/10.1080/08832323.2015.1122562

Britton, E., Simper, N., Leger, A. & Stephenson, J. (2017). Assessing teamwork in undergraduate education: A measurement tool to evaluate individual teamwork skills. *Assessment & Evaluation in Higher Education*, 42(3), 378-397. https://doi.org/10.1080/02602938.2015.1116497

Chyung, S. Y., Roberts, K., Swanson, I. & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, 56(10), 15-23. https://doi.org/10.1002/pfi.21727 [also https://scholarworks.boisestate.edu/ipt_facpubs/87/]

Coll, R. K. & Zegwaard, K. E. (2006). Perceptions of desirable graduate competencies for science and technology new graduates. *Research in Science & Technological Education*, 24(1), 29-58. https://doi.org/10.1080/02635140500485340

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C. & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, 7(3), 179-199.

Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. https://doi.org/10.1080/03075079912331379935

Druskat, V. U. & Pescosolido, A. T. (2002). The content of effective teamwork mental models in self-managing teams: Ownership, learning and heedful interrelating. *Human Relations*, 55(3), 283-314. https://doi.org/10.1177/0018726702553001

Ellis, A. P., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R. & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology*, 58(3), 641-672. http://digitalcommons.ilr.cornell.edu/articles/46/

Falchikov, N. & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. https://doi.org/10.3102/00346543070003287

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J. & Conde, M. Á. (2015). Using learning analytics to improve teamwork assessment. *Computers in Human Behavior*, 47, 149-156. https://doi.org/10.1016/j.chb.2014.11.050

Hastie, C., Fahy, K. & Parratt, J. (2014). The development of a rubric for peer assessment of individual teamwork skills in undergraduate midwifery students. *Women and Birth*, 27(3), 220-226. https://doi.org/10.1016/j.wombi.2014.06.003

Hastie, C. R. (2018). 'TeamUP': An approach to developing teamwork skills in undergraduate midwifery students. *Midwifery*, 58, 93-95. https://doi.org/10.1016/j.midw.2017.12.026

Horvath, L. & Tobin, T. J. (2001). Twenty-first century teamwork: Defining competencies for virtual teams. In M. M. Beyerlein, D. A. Johnson & S. T. Beyerlein (Eds.), *Virtual teams* (pp. 239-258): Emerald Group Publishing Limited. https://www.emeraldinsight.com/doi/abs/10.1016/S1572-0977%2801%2908016-5

Keebler, J. R., Dietz, A. S., Lazzara, E. H., Benishek, L. E., Almeida, S. A., Toor, P. A., King, H. B. & Salas, E. (2014). Validation of a teamwork perceptions measure to increase patient safety. *BMJ Quality & Safety*, 23(9), 718-726. https://doi.org/10.1136/bmjqs-2013-001942

LaBeouf, J. P., Griffith, J. C. & Roberts, D. L. (2016). Faculty and student issues with group work: What is problematic with college group assignments and why? *Journal of Education and Human Development*, 5(1), 13-23. https://doi.org/10.15640/jehd.v5n1a2

Leggat, S. G. (2007). Effective healthcare teams require effective team members: Defining teamwork competencies. *BMC Health Services Research*, 7(1), 17. https://doi.org/10.1186/1472-6963-7-17

Liu, N.-F. & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290. https://doi.org/10.1080/13562510600680582

Mickan, S. & Rodger, S. (2000). Characteristics of effective teams: A literature review. *Australian Health Review*, 23(3), 201-208. https://doi.org/10.1071/AH000201

Nadler, J. T., Weston, R. & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, 142(2), 71-89. https://doi.org/10.1080/00221309.2014.994590

Parratt, J. A., Fahy, K. M., Hutchinson, M., Lohmann, G., Hastie, C. R., Chaseling, M. & O'Brien, K. (2016). Expert validation of a teamwork assessment rubric: A modified Delphi study. *Nurse Education Today*, 36, 77-85. http://doi.org/10.1016/j.nedt.2015.07.023

Quek, A.-H. (2005). Learning for the workplace: A case study in graduate employees' generic competencies. *Journal of Workplace Learning*, 17(4), 231-242. https://doi.org/10.1108/13665620510597185

R Core Team. (2016). *R: A language and environment for statistical computing. Version 3.3.0*. http://www.R-project.org/

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14(1), 57-74. https://doi.org/10.1207/s15327906mbr1401_4

Revelle, W. (2016). *psych: Procedures for Personality and Psychological Research*. Version 1.6.4. https://CRAN.R-project.org/package=psych

Revelle, W. & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145-154. https://doi.org/10.1007/s11336-008-9102-z

Rosen, M. A., Salas, E., Wilson, K. A., King, H. B., Salisbury, M., Augenstein, J. S., Robinson, D. & Birnbach, D. J. (2008). Measuring team performance in simulation-based training: Adopting best practices for healthcare. *Simulation in Healthcare*, 3(1), 33-41. https://doi.org/10.1097/SIH.0b013e3181626276

Rousseau, V., Aubé, C. & Savoie, A. (2006). Teamwork behaviors: A review and an integration of frameworks. *Small Group Research*, 37(5), 540-570. https://doi.org/10.1177/1046496406293125

Ruiz Ulloa, B. C. & Adams, S. G. (2004). Attitude toward teamwork and effective teaming. *Team Performance Management: An International Journal*, 10(7/8), 145-151. https://doi.org/10.1108/13527590410569869

Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, 52(1), 471-499. https://doi.org/10.1146/annurev.psych.52.1.471

Sijtsma, K. & van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, 38(4), 505-528. http://doi.org/10.1207/s15327906mbr3804_4

Sijtsma, K. & van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology*, 70(1), 137-158. https://doi.org/10.1111/bmsp.12078

Sinche, M., Layton, R. L., Brandt, P. D., O'Connell, A. B., Hall, J. D., Freeman, A. M., . Harrell, J. R., Jeanette Gowen Cook , Patrick J. Brennwald Brennwald, P. (2017). An evidence-based evaluation of transferrable skills and job satisfaction for science PhDs. *PLoS ONE*, 12(9), e0186023. https://doi.org/10.1371/journal.pone.0185023

Stochl, J., Jones, P. B. & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*, 12(74). https://doi.org/10.1186/1471-2288-12-74

Taylor, C., Munro, A. J., Glynne-Jones, R., Griffith, C., Trevatt, P., Richards, M. & Ramirez, A. (2019). Multidisciplinary team working in cancer: What is the evidence? *British Medical Journal*, 340, c951. https://doi.org/10.1136/bmj.c951

Valentine, M. A., Nembhard, I. M. & Edmondson, A. C. (2015). Measuring teamwork in health care settings. A review of survey instruments. *Medical Care*, 53(4), e16-e30. https://doi.org/10.1097/MLR.0b013e31827feef6

Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, 48(5), 1-27. https://doi.org/10.18637/jss.v048.i05

Van Ginkel, J. R., van der Ark, L. A. & Sijtsma, K. (2007). Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. *Multivariate Behavioral Research*, 42(2), 387-414. https://doi.org/10.1080/00273170701360803

Van Zundert, M., Sluijsmans, D. & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Varela, O. & Mead, E. (2018). Teamwork skill assessment: Development of a measure for academia. *Journal of Education for Business*, 93(4), 172-182. https://doi.org/10.1080/08832323.2018.1433124

Wind, S. A. (2017). An instructional module on Mokken scale analysis. *Educational Measurement: Issues and Practice*, 36(2), 50-66. https://doi.org/10.1111/emip.12153

Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123-133. https://doi.org/10.1007/s11336-003-0974-7

Zinbarg, R. E., Yovel, I., Revelle, W. & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement*, 30(2), 121-144.

Appendix 1

Providing peer feedback

You are required to complete one TeamUP rubric (survey) for yourself and one for each of your teammates. Please follow the instructions on Blackboard about submission of these documents.

There are 5 domains in total, each with key skills that are important for effective group work. For each skill, there are 7 options to rate yourself and each team member.

You are also invited to add a thoughtful and constructive comment aimed to help your teammate develop their teamwork skills.

For each team member, consider the following question using the scale below:

How well did the individual student demonstrate each particular skill?

1	2	3	4	5	6	7
Unacceptable behaviour or demonstrated minimal or no demonstration of skills, major improvements are needed		To a limited extent, a **below average** team member, moderate improvements needed		To a moderate extent, **an average** team member, some minor improvements needed		Mastery, well **above average** team member, no improvement needed

Your name:	
Name of team member being assessed:	

Domain 1 - Planning	The actions indicate the team member has worked in synergy with others to plan a high quality project							
a. Electing and supporting a project manager	1	2	3	4	5	6	7	NA
b. Defining and agreeing on team goals and objectives	1	2	3	4	5	6	7	NA
c. Defining and agreeing on quality standards for each part of the plan	1	2	3	4	5	6	7	NA
d. Contributing to the development of the plan	1	2	3	4	5	6	7	NA
e. Setting and agreeing realistic timeframes for each part of the plan	1	2	3	4	5	6	7	NA
f. Participating in role allocations based on individual skills and learning needs	1	2	3	4	5	6	7	NA
g. Willingly taking on a team role that can be completed on time to a quality standard	1	2	3	4	5	6	7	NA
Domain 2 - Environment	These actions fostering a sense of trust and inclusiveness for each team member							
a. Exhibiting an open, gentle, polite and friendly manner	1	2	3	4	5	6	7	NA
b. Demonstrating self-awareness and emotional regulation	1	2	3	4	5	6	7	NA
c. Demonstrating sensitive awareness of the feelings of others (including interpreting body language)	1	2	3	4	5	6	7	NA
d. Actively contributing to team discussions	1	2	3	4	5	6	7	NA
e. Cooperating with others to achieve project goals	1	2	3	4	5	6	7	NA

f.	Following up with others when there is concern about their feelings or contribution	1	2	3	4	5	6	7	NA
g.	Showing respect for the contributions of others (even if in disagreement)	1	2	3	4	5	6	7	NA
h.	Expressing genuine gratitude and praise generously	1	2	3	4	5	6	7	NA
Domain 3 - Facilitating the contributions of others		These actions preventing, recognising and/or addressing conflict in ways that strengthen overall team cohesiveness and effectiveness							
a.	Taking turns at leading/coordinating a team meeting (including agenda preparation)	1	2	3	4	5	6	7	NA
b.	Taking turns at keeping and distributing brief meeting minutes (with action items and deadlines)	1	2	3	4	5	6	7	NA
c.	Leading and/or participating in teambuilding processes	1	2	3	4	5	6	7	NA
d.	Establishing and honouring team ground rules	1	2	3	4	5	6	7	NA
e.	Ensuring that decisions are made in a timely manner	1	2	3	4	5	6	7	NA
f.	Listening attentively without interrupting and raising hand to speak	1	2	3	4	5	6	7	NA
g.	Participating in consensus-building decision-making	1	2	3	4	5	6	7	NA
h.	Inviting other team members to contribute	1	2	3	4	5	6	7	NA
Domain 4 - Managing conflict		These actions demonstrating that the team member has made a high-quality, individual contribution to the team project							
a.	Being appropriately assertive: neither dominating, submissive, nor passive aggressive	1	2	3	4	5	6	7	NA
b.	Expressing concerns with team/team members in a constructive manner	1	2	3	4	5	6	7	NA
c.	Minimising unnecessary conflict by project planning and management	1	2	3	4	5	6	7	NA
d.	Completing assigned responsibilities on time	1	2	3	4	5	6	7	NA
e.	Participating in the team conflict transformation processes	1	2	3	4	5	6	7	NA
f.	Assisting the team to stay focused on the overall team goal	1	2	3	4	5	6	7	NA
g.	Approaching conflict with the aim to de-escalate	1	2	3	4	5	6	7	NA
h.	Being open to receiving and reflecting upon criticism of self	1	2	3	4	5	6	7	NA

i.	Challenging team processes that are not conducive to the achievement of team goals	1	2	3	4	5	6	7	NA
Domain 5 - Contributing to team project		These actions ensuring the process of team interactions are effective in progressing the project plan							
a.	Demonstrating sufficient technological skills	1	2	3	4	5	6	7	NA
b.	Demonstrating relevant content knowledge	1	2	3	4	5	6	7	NA
c.	Adhering to academic standards for writing	1	2	3	4	5	6	7	NA
d.	Submitting assigned work at the agreed quality standard	1	2	3	4	5	6	7	NA
e.	Submitting assigned work within the agreed timeframe	1	2	3	4	5	6	7	NA
f.	Appropriately critiquing the work of others	1	2	3	4	5	6	7	NA
g.	Working to integrate the output of team members into the project	1	2	3	4	5	6	7	NA
h.	Evaluating the quality of the whole project and making needed changes	1	2	3	4	5	6	7	NA

Brett Vaughan is a lecturer in clinical education in the Department of Medical Education, University of Melbourne, Australia and a Professional Fellow in the School of Health & Human Sciences at Southern Cross University, New South Wales, Australia. His interests centre on assessment and evaluation in health professions education.
Email: brett.vaughan@unimelb.edu.au
Web: https://www.findanexpert.unimelb.edu.au/display/person824488#tab-publications

Jacqui Yoxall is a registered psychologist and Associate Professor and Group Lead of Allied Health and Midwifery, in the School of Health and Human Sciences, Southern Cross University, New South Wales, Australia. Her clinical and research interests centre on assessment of human behaviour.
Email: jacqui.yoxall@scu.edu.au

Sandra Grace is a health services researcher and an Associate Professor in the School of Health and Human Sciences, Southern Cross University. Her research interests include integrative and complementary medicine, models of healthcare and interprofessional education and practice.
Email: sandra.grace@scu.edu.au

Please cite as: Vaughan, B., Yoxall, J. & Grace, S. (2019). Peer assessment of teamwork in group projects: Evaluation of a rubric. *Issues in Educational Research*, 29(3), 961-978. http://www.iier.org.au/iier29/vaughan.pdf